

Sparse Linear Models applied to Power Quality Disturbance Classification

Andrés F. López-Lopera, Mauricio A. Álvarez, and Ávaro A. Orozco.
 Faculty of Engineering, Universidad Tecnológica de Pereira, Colombia, 660003.
 November 24, 2015

Abstract

Power quality (PQ) analysis describes the non-pure electric signals that are usually present in electric power systems. The automatic recognition of PQ disturbances can be seen as a pattern recognition problem, in which different types of waveform distortion are differentiated based on their features. Similar to other quasi-stationary signals, PQ disturbances can be decomposed into time-frequency dependent components by using time-frequency or time-scale transforms, also known as dictionaries. These dictionaries are used in the feature extraction step in pattern recognition systems. Short-time Fourier, Wavelets and Stockwell transforms are some of the most common dictionaries used in the PQ community, aiming to achieve a better signal representation. To the best of our knowledge, previous works about PQ disturbance classification have been restricted to the use of one among several available dictionaries. Taking advantage of the theory behind sparse linear models (SLM), we introduce a sparse method for PQ representation, starting from overcomplete dictionaries. In particular, we apply Group Lasso. We employ different types of time-frequency (or time-scale) dictionaries to characterize the PQ disturbances, and evaluate their performance under different pattern recognition algorithms. We show that the SLM reduce the PQ classification complexity promoting sparse basis selection, and improving the classification accuracy.

1 Introduction

Power quality (PQ) and reliability, and equipment diagnostics, control and protection are some of the case studies of signal processing in electric power systems [1]. PQ studies evaluate the waveform of voltage and current signals respect to pure sinusoidal waveforms with a single frequency component (e.g. 50 or 60 Hz) [2, 3]. Non-linear electronic devices such as cycloconverters, alternators, and current starters [3, 4] introduce undesirable harmonic distortions to the network. Other examples of sources that distort pure sinusoidal electric signals include electrical loads with high values, electrical faults on the network [3], tap changer [5], and capacitors switching banks [4]. Depending on the source of PQ disturbances, we can find distortions such as swell/sag, harmonics, flickers, notching and transients [1–3].

The automatic recognition of PQ disturbances is an important research field in PQ analysis. It can be stated as a pattern recognition problem, in which the different types of waveform distortions are differentiated based on their features [6]. For the feature extraction, signal processing practitioners in PQ, have employed several basis function superposition, also known as dictionaries [7]. Similar to other quasi-stationary signals, PQ disturbances can be decomposed into time-frequency dependent components by using time-frequency or time-scale dictionaries [1, 6]. Short time Fourier transform (STFT) [8], Wavelets transform (WT) [9, 10], and S-transform (ST) [11, 12] are commonly used in PQ representation for the feature extraction. After the feature extraction step, next stage is the automatic pattern recognition for the classification problem. Classifiers based on K-nearest neighbours (K-NN) [13], Bayesian algorithms [14], support vector machines (SVM) [15], and artificial neural networks (ANN) [16] have been used for PQ pattern classification.

Several researches in PQ disturbance classification have used different combinations of dictionaries and classifiers to improve the PQ classification accuracy for PQ disturbances [6]. Gargoom et al. [17] used a K-NN classifier with statistical features computed either from a WT, a ST or a Hilbert-Huang transform (HHT), obtaining a classification accuracy around 82 ~ 89%. In [11] and [18], the authors employed a ST to identify short durations disturbances, with an ANN classifier, obtaining a performance between 80 ~ 90% for synthetic generated signals with 30-dB noise level. Similarly, Jayasree et al. used the energy of intrinsic mode functions (IMF) components computed from a HHT as features for an ANN classifier [19], obtaining 95 ~ 98%. Most recently, in [10, 16, 20], the authors have used the detailed coefficients from a WT, or some of its variant such as Wavelet multi-resolution analysis, together with classifiers based on either a SVM or an ANN. They obtained a performance between 95 ~ 98%.

To the best of our knowledge, previous works in PQ disturbance representation and classification have been limited to the use of single dictionaries for feature extraction step. This is, they only use either GT, HHT, WT or ST together with some classifiers [8–12]. However, research in signal processing has shown that combining several dictionaries for signal analysis

can improve the robustness and numerical stability in the feature extraction step [7, 21, 22]. Combinations of complete dictionaries for signal representations are also known as overcomplete representations (OR), and they are useful a) for increasing the richness of the representation by removing uncertainty when choosing the “proper” dictionary [21], b) for allowing a non-unique representation with the possibility of adaptation [7], c) for increasing the robustness in presence of noise, and d) for increasing the flexibility for matching any structure present in the data [22]. However, OR increase the length of the coefficient vector of representation extracted from the signals, and may include information that is not relevant or that may be confusing for the classification step.

In order to remove redundant information, before the classification step, research communities in statistical signal processing, statistics and machine learning have proposed different methods for automatically performing basis selection in linear models. These methods are known as ℓ_1 regularization methods or sparse linear models (SLM) [23]. SLM highlight the relevant features in signal representation, making zero (or approximately zero) the contribution of less relevant ones. This property is known as sparsity [24]. Taking advantage of the sparsity due to SLM, it is possible to build OR avoiding the redundant information among the dictionaries. This approach tends to increase the PQ representation accuracy due to OR, and reduces PQ classification complexity due to sparse tendency. Since our OR approach for PQ disturbances requires the analysis of variables (coefficient of representations) that are grouped per each dictionary (e.g. GT, WT or ST), in this paper we employ a SLM known as Group Lasso for PQ disturbance classification. We evaluate the performance of Group Lasso using different classifiers with either one dictionary at a time or with OR. We employ the discrete-time form of the GT, WT and ST dictionaries. Finally, for the PQ classification step, we use different statistical classifiers, namely, K-NN, SVM, ANN, and two Bayesian classifiers based on linear (LDC) and quadratic (QDC) discriminant functions.

This paper is organized as follows. Materials and methods are described in section 2. In section 3, we compare and discuss the results. Finally, section 4 shows the conclusions.

2 Materials and Methods

2.1 Power quality disturbances

PQ determines the fitness of electric power to consumer devices, evaluating the quality of voltage and current signals. We will use two concepts to describe PQ disturbances, namely, events and variations. Events are sudden distortions which occur in specific time intervals. On the other hand, variations are steady-state or quasi-steady-state disturbances which require continuous measurements [1]. This paper focusses on four types of disturbances: voltage or current variations (e.g. swell, sag and flicker), harmonic distortions, notching and transient events (e.g. impulsive and oscillatory).

2.1.1 Swell and sag events

swells describe an increase between 1.10 and 1.80 p.u. (per unit) over the RMS (Root Mean Square) value in an electric signal during an interval between 10 ms and 1 minute (short time), or over 1 minute (long time) [25]. Opposite to swells, sags describe a decrease between 0.10 and 0.90 p.u. below the RMS value. This kind of distortions are commonly produced by tap charger, electrical loads with high values and electrical faults on the network [3].

2.1.2 Flicker variations

fluctuations between $\pm 0.1\%$ and $\pm 7\%$ respect to the RMS value with a frequency lower than 25 Hz. Cycloconverters, electric arc furnace and high current starters are the main sources of flicker effects [3, 4].

2.1.3 Harmonic variations

distortions which present different frequency components in electric signals. These components correspond to integer multiples of the fundamental frequency [2]. Harmonics appear usually by the connection of non-linear devices to the network [4].

2.1.4 Notching variations

they introduce high harmonic and non-harmonic frequencies in higher voltage systems, due to the normal operation of electronic devices when current is commutated from one phase to another [3, 5].

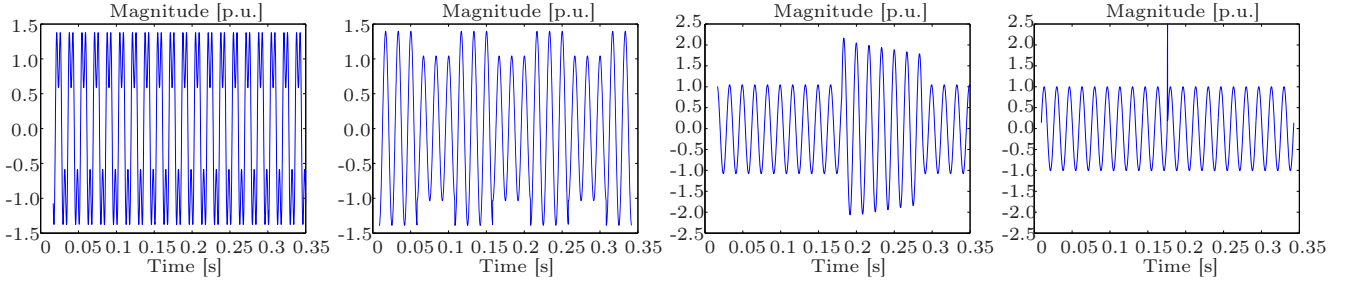


Figure 1: PQ disturbance examples. Figure shows different PQ disturbances such as harmonic distortion, flicker variation, swell event, and impulsive transient (from left to right). The examples correspond to the synthetic dataset described in section 2.6

2.1.5 Transient events

they can be classified in two types, impulsive and oscillatory events. The impulsive transients are unidirectional events during a short time lapse between 50 ns and 1 ms. They are produced by electric discharges or inductive charge commutation (motors) [3, 4]. Oscillatory transients are bidirectional events during an interval time between 0.3 ms and 50 ms. The oscillations occur commonly by power factor supplies using capacitors switching banks [3, 4].

Table 1 summarizes the characteristics of the PQ disturbances listed above. Figure 1 shows examples of PQ distortions from the synthetic dataset described in section 2.6.

Table 1: Summary of PQ disturbances.

Disturbance	Spectral Content	Duration	Magnitude [p.u.]
Events			
1. Swells			
1.1. Short time		10 ms-1 min	1.1-1.8
2.1. Long time		> 1 min	1.1-1.8
2. Sags			
1.2. Short time		10 ms-1 min	0.1-0.9
2.2. Long time		> 1 min	0.1-0.9
3. Transients			
3.1. Impulsive	5 ns rise	< 50 ns	
3.2. Oscillatory			
3.2.1. Low frequency	< 5 KHz	0.3-50 ms	0-4
3.2.2. Mid frequency	5-500 KHz	20 ms	0-8
3.2.3. High frequency	0.5-5 MHz	5 ms	0-4
Variations			
4. Harmonic Distortion	k=2-40 (th)	steady state	< 0.2
5. Flicker	< 25 Hz	intermittent	0.001-0.07
6. Notching		steady state	

Spectral Content and Duration correspond to the typical frequencies and time lapses. Magnitude represents the variation in amplitude and it is given in per unit (p.u.).

2.2 Overcomplete representation

Be $\mathbf{y} \in \mathbb{R}^N$ a vector which represents a discrete-time signal (e.g. a PQ disturbance) of length N . A signal representation using a superposition of atoms (basis functions) $\phi_\gamma(t)$ is known as a complete representation [26]. Here, the collection of parametrized atoms $\Phi = (\phi_\gamma : \gamma \in \Gamma)$ is known as a dictionary, where the parameter γ has different interpretations according to the indexation of the variables: frequency, time-frequency or time-scale values. In general, it is possible to assume that \mathbf{y} can be expressed by the following linear matrix equation

$$\mathbf{y} = \Phi\boldsymbol{\beta}, \quad (1)$$

where $\boldsymbol{\beta} \in \mathbb{R}^M$ and $\Phi \in \mathbb{R}^{N \times M}$ are the vector of coefficients and the dictionary related to the representation, respectively. An overcomplete representation (OR) is obtained by the combination of several dictionaries in order to represent completely the signal \mathbf{y} , obtaining a new dictionary given by $\Psi = (\Phi_v : v \in \Upsilon)$ [7]. In this paper, we analyse the PQ disturbances in terms of the vector of coefficients $\boldsymbol{\beta}$ (analysis step). These coefficients are subsequently used for feature extraction, to feed the PQ disturbance classifiers. We also use $\boldsymbol{\beta}$ to reconstruct or synthesize (synthesis step) the PQ disturbances vector computed from the analysis step [7]. We evaluate the quality of the reconstruction provided by the sparse vector obtained after applying Group Lasso.

2.3 Sparse linear models for grouped variable selection: Group Lasso

Tibshirani [27] proposes a method for estimation in linear models known as Lasso (least absolute shrinkage and selection operator). To solve the inverse problem from the Equation 1, he demonstrated that Lasso obtains a sparse representation due to the minimization of the cost function given by

$$\hat{\beta}_\lambda = \arg \min_{\beta} \left\{ \frac{1}{2} \|\mathbf{y} - \Phi\beta\|_2^2 + \lambda \|\beta\|_1 \right\},$$

where $\hat{\beta}_\lambda \in \mathbb{R}^M$ is the vector of estimated coefficients which depends of the regularization parameter λ , $\|\cdot\|_2$ and $\|\cdot\|_1$ correspond to the ℓ_2 -norm, and ℓ_1 -norm, respectively. Here, Lasso assumes that there is a direct and unique correspondence between parameters and variables (1:1 correspondence), performing the variable selection individually. However, the individual selection of variables produces a not satisfactory solution for grouped variables (e.g. coefficients β_v related to bases Φ_v from the dictionary Ψ) [28]. To deal with grouped variable estimation, an extension of Lasso was proposed in [29]. This method is known as Group Lasso. Group Lasso assumes that the vector β is partitioned in G groups where the penalty is an intermediate between ℓ_1 and ℓ_2 regularizations. Meier et al. [28] showed that Group Lasso has the attractive property it performs the variable selection at the group level, promoting sparsity over $\hat{\beta}_\lambda$ for large values of λ . For linear regression, the cost function for Group Lasso is given by

$$\hat{\beta}_\lambda = \arg \min_{\beta} \left\{ \frac{1}{2} \|\mathbf{y} - \Phi\beta\|_2^2 + \lambda \sum_{g=1}^G \|\beta_g\|_{\mathbf{K}_g} \right\}, \quad (2)$$

with $\|\beta_g\|_{\mathbf{K}_g} = (\beta_g^\top \mathbf{K}_g \beta_g)^{1/2}$. Yuag et al. [29] implemented a Group Lasso approach of the shooting algorithm proposed in [30], assuming $\mathbf{K}_g = p_g I_{p_g}$ with I_{p_g} an identity matrix which depends on the length of g -th group. According to this proposition, the necessary and sufficient condition for $\beta = [\beta_1^\top, \dots, \beta_G^\top]^\top$ to be a solution of Equation (2) is given by

$$\beta_g = \left[1 - \frac{\lambda \sqrt{p_g}}{\|\mathbf{S}_g\|} \right] \mathbf{S}_g, \quad (3)$$

where $\mathbf{S}_g = \Phi_g^\top (\mathbf{y} - \Phi \beta_{-g})$, with $\beta_{-g} = [\beta_1^\top, \dots, \beta_{g-1}^\top, \mathbf{0}, \beta_{g+1}^\top, \dots, \beta_G^\top]^\top$. The solution of the Equation (2) can therefore be obtained by iteratively applying the Equation (3) for $g = 1, 2, \dots, G$ [29]. In this paper, each group corresponds to a single dictionary, and β_g are the representation coefficients related to the g -th dictionary.

2.4 Time-frequency and time-scale dictionaries

2.4.1 Short Time Fourier Transform (STFT)

STFT is the general expression of time-frequency dictionaries. The STFT uses the concept of a window function to compute the Fourier transform in a specific interval of the analysis signal. The STFT of a discrete-time series $y[n]$ is giving by the expression

$$\text{STFT}\{y[n]\} = \sum_{n=-\infty}^{\infty} y[n] h[n - m_l] \exp\{-j\omega_k[n - m_l]\},$$

where $h[n - m_l]$ is the window function with location m_l , and $f_k = \omega_k/2\pi$ are the frequency terms which we use for the representation. In PQ disturbance, the Gabor transform (GT) is the most common STFT where $h[\cdot]$ is defined by a Gaussian function [8]. Other types of window functions used in PQ analysis are the Hamming function and the Hann function [8]. For real-valued representations, the atoms $\phi_\gamma[n]$ for GT dictionary follow

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-(n - m_l)^2/2\sigma^2\right\} \cdot \cos(2\pi f_k(n - m_l)),$$

where $\gamma = (f_k, m_l)$, $f_k = kf_o$ ($k = 1, 2, \dots, K$), and $m_l = lm_o$ ($l = 1, 2, \dots, L$). Here, K (total frequencies), L (total locations), f_o (fundamental frequency) and m_o (fundamental location) are constants. The main disadvantage of GT in PQ disturbances is related to its resolution for signal representation. A higher resolution in time leads to a lower resolution in frequency, and vice versa [8]. In addition, we need to fix the length scale σ [9].

2.4.2 Wavelet Transform (WT)

WT is other example of time-frequency transforms applied to PQ disturbances [10]. Similarly to STFT, the WT uses a time-scale dictionary characterized by irregular functions called wavelets. Contrary to STFT, WT gives the possibility to vary the scale factor, obtaining higher resolutions in time when it is required. Here, we can fix the length of the wavelet to detect high frequencies [9]. WT is described by

$$\text{CWT}\{y[n]\} = \frac{1}{\sqrt{\sigma_\nu}} \sum_{n=-\infty}^{\infty} y[n] \psi \left[\frac{n - m_l}{\sigma_\nu} \right],$$

where $\psi[(n - m_l)/\sigma_\nu]$ is the wavelet function with locations m_l and length-scale σ_ν . Mexican Hat Wavelet Transform (MHWT), Daubechies, Morlets and Symlets wavelets have been applied to PQ disturbance representation [10]. In this paper, we use the MHWT. Its atoms $\phi_\gamma[n]$ can be expressed by the following form

$$\frac{2}{\sqrt{3\sigma_\nu} \pi^{1/4}} [1 - (n - m_l)^2 / \sigma_\nu^2] \cdot \exp\{-(n - m_l)^2 / 2\sigma_\nu^2\},$$

where $\gamma = (m_l, \sigma_\nu)$, $\sigma_\nu = \nu \sigma_o$ ($\nu = 1, 2, \dots, V$), and $m_l = l m_o$. Here, V (total scales), L , σ_o (principal scale) and m_o are constants. The WT has been used to characterize PQ disturbances such as voltage variation (e.g. swells and sags) [10], and transients distortions (e.g. impulsive and oscillatory), due to its variable length-scale [9].

2.4.3 Stockwell Transform (ST)

ST is a generalization of the STFT, extending the continuous WT and overcoming some of its disadvantages. ST includes a correction factor which gives information about phase contribution in the frequency domain, improving the performance for PQ representations. The phase correction can characterize the real and the imaginary spectrum components independently [11]. The ST combines the time-frequency and time-scale concepts, given

$$\text{ST}\{y[n]\} = \sum_{n=-\infty}^{\infty} y[n] g[n - m_l, \sigma_\nu] \exp\{-j\omega_k(n - m_l)\},$$

where $g[n - m_l, \sigma]$ is a Gaussian window with location m_l and length-scale σ_ν . The atoms $\phi_\gamma[n]$ for ST are given by

$$\frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\{-(n - m_l)^2 / 2\sigma_\nu^2\} \cdot \cos(2\pi f_k(n - m_l)),$$

where $\gamma = (f_k, m_l, \sigma_\nu)$, $f_k = k f_o$, $m_l = l m_o$, and $\sigma_\nu = \nu \sigma_o$. ST shows a high performance in PQ representations for several types of distortions such as swell, sag and flickers [11, 12].

In general, it is possible to express many dictionaries in a matrix form $\Phi = (\phi_\gamma : \gamma \in \Gamma)$. Also, we can combine several dictionaries to achieve a new and most robust dictionary by $\Psi = (\Phi_v : v \in \Upsilon)$, where γ and v are indexation parameters. It is noteworthy that these dictionaries do not need to be orthogonal to obtain an OR [7].

2.5 Classifiers for automatic recognition

PQ disturbance classification can be seen as a pattern recognition problem, in which each class corresponds to a type of PQ disturbance. We start with a set $\{\mathbf{x}_p, \ell_p\}_{p=1}^P$, where $\mathbf{x}_p \in \mathbb{R}^D$ is a feature vector extracted from the signal p (we will see later how the elements in \mathbf{x}_p are computed from the vector of coefficients β_p), and ℓ_p is the corresponding label or class for signal p . P is the total number of samples. The automatic recognition of disturbances is made of two stages: training and testing stages. In the first stage, the classifiers use a set of samples $\{\mathbf{x}_t, \ell_t\}_{t=1}^T$, to learn a mapping between the input space and the different PQ disturbances. In the second stage, the classifiers use the information obtained from the training set to predict the labels of a new set of samples $\{\mathbf{x}_u\}_{u=1}^U$. Notice that T and U correspond to the number of samples for the training and test datasets, respectively. In this subsection, we give a brief description of the classifiers we used to validate the performance of SLM in the PQ classification step. The classifiers that we use are state-of-the-art, and the theory behind them can be found in deep in any pattern recognition or machine learning textbook [23, 24].

2.5.1 K-nearest neighbours (K-NN)

K-NN is a non-parametric method where the labels of the K nearest points in the training set are taken into account to predict the labels for the test set [13]. Here, the proximity between neighbours is defined by the Euclidean distance [23].

Table 2: Parametric equation for PQ disturbance simulations.

PQ disturbance	Notation	Controlling Parameters	Parametric Equation
Pure Signal	—	—	$y(t) = \sin(\omega t)$
Harmonic Distortion	C_1	$0.05 \leq \alpha_k \leq 0.15, \sum_k \alpha_k^2 = 1$	$y(t) = \sum_k \alpha_k \sin(k\omega t) \text{ con } k = 3, 5, 7, 9.$
Swell	C_2	$0.1 \leq \alpha \leq 0.8, 6 \leq t_2 - t_1 \leq 10T$	$y(t) = [1 + \alpha\{u(t_2) - u(t_1)\}] \sin(\omega t)$
Sag	C_3	$0.1 \leq \alpha \leq 0.8, 6 \leq t_2 - t_1 \leq 10T$	$y(t) = [1 - \alpha\{u(t_2) - u(t_1)\}] \sin(\omega t)$
Flicker Effect	C_4	$0.1 \leq \alpha \leq 0.2, 5 \leq f_f \leq 25\text{Hz}$	$y(t) = [1 + \alpha \sin(2\pi f_f t)] \sin(\omega t)$
Notch	C_5	$0.1 \leq \alpha_k \leq 0.4, 0.01T \leq \Delta t_k \leq 0.05T$	$y(t) = \sin(\omega t) - \text{sign}\{\sin(\omega t)\} \sum_k \alpha_k [u(t_k + \Delta t_k/2) - u(t_k - \Delta t_k/2)]$
Impulsive Transient	C_6	$1.5 \leq \alpha \leq 2.5, 0.01T \leq t_2 - t_1 \leq 0.02T$	$y(t) = \sin(\omega t) + \alpha \text{sign}\{\sin(\omega t)\} [u(t_2) - u(t_1)]$
Oscillatory Transient	C_7	$0.1 \leq \alpha \leq 0.9, 0.01T \leq t_2 - t_1 \leq 0.03T$ $0.1 \leq \beta \leq 0.2, 5 \leq k \leq 80$	$y(t) = \sin(\omega t) + \alpha \exp\{-t/\beta\} \sin(k\omega t) [u(t_2) - u(t_1)]$

T and ω are the period and the angular frequency, respectively. $u\{\cdot\}$ and $\text{sign}\{\cdot\}$ are the unit step and the sign functions, respectively.

2.5.2 Bayesian classifiers

there are two types of non-parametric Bayesian classifiers commonly used for patter recognition, namely, the linear (LDC) and the quadratic (QDC) discriminant classifiers. LDC finds a linear combination of features which is used as a linear classifier. This method assumes normal densities with equal covariance for each class, and the joint covariance is given by the weighted average of the class covariances. QDC also assumes that measurements for each class are normally distributed, but there is not the assumption that covariances for each class are identical. Both methods consist of the computation of the sample means and the sample covariances of each class [23, 24].

2.5.3 Support vector machines (SVM)

SVM are decision machines that maximize the separating margin between two classes given a set of training data [24]. The risk function of SVM involves the use of Lagrange multipliers, a set of Karush-Kuhn-Tucker conditions, and variables to control the trade-off between the slack variable penalty and a boundary that separate both classes [23]. The SVM maximize the Lagrangian dual risk function which involves inner products between training data, which later are replaced for kernel functions. Some common kernels for PQ disturbance classification are the exponential, and the Gaussian kernels (or radial basis function, RBF) [15] [31].

2.5.4 Artificial neural networks (ANN)

ANN use non-linear activation functions over the outputs of linear models which can be adjusted during the training stage [23]. More details can be found in [23]. The most common classifiers for PQ disturbance are based on ANN and the feed-forward back-propagation scheme [11, 16].

2.6 Procedure

2.6.1 Dataset

Table 2 shows the parametric equations used to construct the synthetic dataset for different types of PQ disturbances, namely, swells, sags, flicker effects, notches, impulsive and oscillatory transients. Those equations are based on [20, 31–33]. We generated 1330 PQ disturbances with 190 samples per each type. Each disturbance has 2101 discrete-time values for an interval between 0 and 0.7 seconds.

2.6.2 Group Lasso

we implement the shooting algorithm for Group Lasso described in section 2.3, using a penalty factor $\lambda = 1 \times 10^{-3}$, and a convergence tolerance equal to 1×10^{-15} . These values were selected manually in order to obtain a signal reconstruction error lower than 1×10^{-3} , and for preserving relevant representation coefficients. Here, each group corresponds to a single dictionary, and β_g are the representation coefficients related to the g -th dictionary. The percentage of sparsity for the coefficients β is computed by the sum of the number of coefficient lower than 1×10^{-4} , and dividing the result by the length of the vector β .

2.6.3 Dictionaries

for each time-frequency and time-scale dictionary, we use a location factor $m_o = 0.005s$. For the GT, we focus in the first 40 harmonics with a length-scale $\sigma = 0.005s$. For MHWT, we used as scale factor $\sigma_o = 0.005s$ with $V = 10$. For ST, we used the first 6 harmonics, and $\sigma_o = 0.005s$ with $V = 6$. However, we add a cosine/sine dictionary for all the cases with the first 40 harmonics, aiming to improve the PQ synthesis results. We denote it as the Harmonics dictionary. Finally, in order evaluate the SLM performance using OR, we combine the GT, MHWT and ST dictionaries which we denote as GWST dictionary.

2.6.4 Feature extraction

after applying the dictionaries proposed in subsection 2.4, we compute the following features from $\{\beta_p\}_{p=1}^P$ to create the feature vectors $\{\mathbf{x}_p\}_{p=1}^P$ for each signal p : mean of the absolute values (F_1), standard deviation (F_2), kurtosis (F_3), Shannon's energy (F_4) and RMS value (F_5). We also include the mean of the absolute values of the derivative β' (F_6)¹. For the PQ disturbance classification step, we normalize the features by subtracting their mean values, and dividing the result using their standard deviation.

2.6.5 Classifiers

for the classifier based on ANN, we use the Neural Network Toolbox provided for Matlab R2013a. We use an ANN made of three hidden layers with 20, 15 and 10 neurons in the first, second, and third layer, respectively, with sigmoid transfer functions [34]. We employ the Levenberg-Marquardt scheme using the MSE criteria for the training stage. For the other classifiers described in 2.5, we use the Matlab Toolbox for Pattern Recognition (PRTools Toolbox²). For the SVM, we use a RBF kernel $k(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2/\sigma^2)$, where σ is known as the bandwidth parameter. The bandwidth parameter, and the regularization parameter for the SVM are tuned by cross-validation. We test the classifiers twenty times with different training sets. We select randomly the 70% of the total samples per each type of PQ disturbance to train the classifiers (133 samples per class), and then we use the other 30% for test stage (57 samples per class). The performance for each test experiment is computed by the sum of the successful classifications of PQ disturbances, and dividing the result using the total number of test samples. Finally, we compute the mean and the standard deviation of the performance obtained in all the experiments.

3 Results and Discussions

To observe the advantages of sparse linear models (SLM) for PQ representation, Figure 2 shows the synthesis and the analysis steps for a single example of a swell event. The swell disturbance used as example appears in Figure (a). The result of the synthesis step is shown for two cases: in (b) without sparsity, and in (c) using Group Lasso. For this result, we use the GWST dictionary which we obtain by combining the GT, MHWT and ST. Both methods can synthesize the PQ distortion in Figure (a), ensuring a low reconstruction error. Moreover, we perform the analysis step in Figures (d) and (e) for the same example. The first, second and third interval separated by the dash line correspond to the coefficients obtained by GT, WT and ST, respectively. To represent the swell example, we see the method without Group Lasso in (d) tends to use all the coefficients from β . However, from Figure (e), Group Lasso achieves the synthesis of the PQ disturbance using slightly less than the 60% of the total coefficients available, selecting a few number of coefficient from each dictionary. Finally, we see that the magnitude of the coefficients obtained with Group Lasso is far smaller compared to the magnitudes obtained without SLM due to the regularization parameter used in Group Lasso.

In order to quantify the level of sparsity produced by Group Lasso over the GWST representation, the synthesis step was performed over all the PQ disturbance examples from the synthetic dataset described in section 2.6. Tables 3 and 4 show the sparsity percentages produced by the method without sparsity, and using Group Lasso, respectively. According to section 2.6.2, we compute the sparsity percentages per each type of PQ disturbance (rows), and each dictionary (columns). Table 3 shows that methods without sparsity tend to use all the coefficient of representation from GWST, making more difficult subsequent PQ studies (e.g. PQ analysis and PQ disturbance classification). According to Table 4, Group Lasso

¹Let $\beta = [\beta_1, \beta_2, \dots, \beta_M]^\top$, the features were calculated as follows

$$F_1 = \frac{1}{M} \sum_{i=1}^M |\beta_i|, \quad F_2 = \left[\frac{1}{M-1} \sum_{i=1}^M (\beta_i - \bar{\beta})^2 \right]^{1/2}, \quad F_3 = \left[\frac{1}{M} \sum_{i=1}^M (\beta_i - \bar{\beta})^4 \right] / \left[\frac{1}{M} \sum_{i=1}^M (\beta_i - \bar{\beta})^2 \right]^2, \quad F_4 = \frac{1}{2} \sum_{i=1}^M \beta_i^2, \\ F_5 = \sqrt{\frac{1}{M} \sum_{i=1}^M |\beta_i|^2}, \quad F_6 = \frac{1}{M} \sum_{i=1}^M |\beta'_i|,$$

where β' is the derivative of β and $\bar{\beta} = \frac{1}{M} \sum_{i=1}^M \beta_i$.

²PRTools Toolbox is available at <http://prtools.org/>.

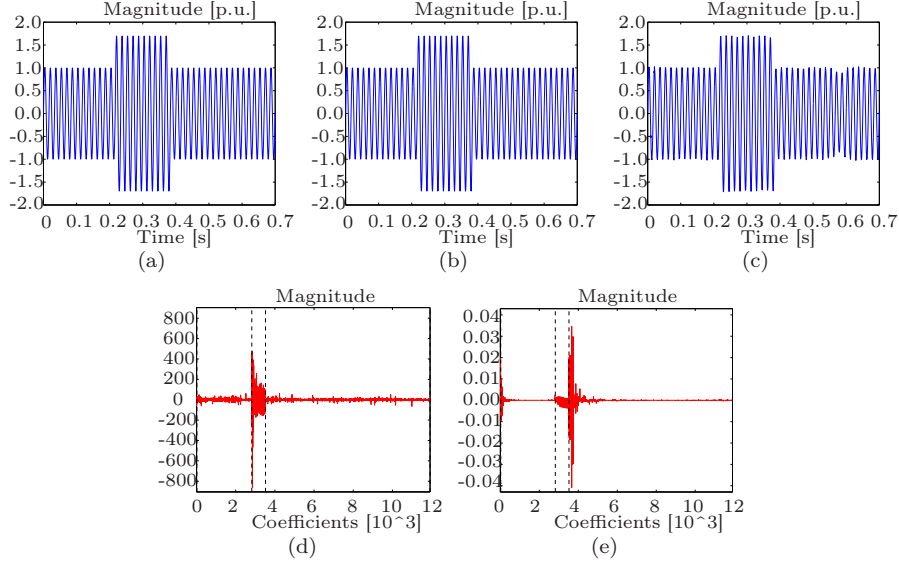


Figure 2: Synthesis and analysis step results for a swell event using GWST dictionary. In (a) we show a swell example. The synthesis step without SLM and using Group Lasso are showed in (b) and (c), respectively. Finally, we show the coefficient β used for synthesis step from Figures (b) and (c), in Figures (d) and (e).

Table 3: Sparsity percentages using GWST without Group Lasso.

Type	Harmonics	GT	MHWT	ST
Harmonic	0	0.0043	0	0.0150
Swell	0	0.0034	0.0001	0.0054
Sag	0	0.0045	0	0.0083
Flicker	0	0.0041	0	0.0057
Notch	0	0	0	0.0015
Impulsive	0	0.0041	0	0.0087
Oscillatory	0	0.0032	0	0.0048

Table 4: Sparsity percentages using GWST with Group Lasso.

Type	Harmonics	GT	MHWT	ST
Harmonic	38.1053	81.0351	10.6900	56.1971
Swell	52.3421	91.0189	7.5048	64.1343
Sag	51.7632	91.0485	7.6622	63.9687
Flicker	61.6579	92.5543	9.5789	65.4090
Notch	19.4737	51.0748	10.8880	32.6568
Impulsive	39.9211	47.8662	8.0383	23.5953
Oscillatory	39.4737	67.7954	10.3418	44.8384

GT: Gabor transform. MHWT: Mexican hat Wavelet transform. ST: Stockwell transforms.

automatically selects the most representative coefficients required to fully represent the PQ disturbances. Also, we notice that the Harmonics, the MHWT and the ST dictionaries contain more relevant components in terms of the disturbance representation, when compared to the GT dictionary.

To evaluate the performance of SLM for PQ classification, we first perform the synthesis step per each disturbance represented through GT, MHWT, ST, and GWST (combining GT, MHWT, ST), obtaining the set of coefficients $\{\beta_p\}_{p=1}^P$. We then apply Group Lasso over the different representations, and compute the feature set $\{\mathbf{x}_p\}_{p=1}^P$ described in 2.6.4. We train the different classifiers, and evaluate their performance over the test set as well as we describe in section 2.6.5

Table 5 shows the PQ classification performance over the test set. When Group Lasso is not applied (first four rows), ST and GWST show better results than GT and WT, independently of the classifier employed. When Group Lasso is applied (last four rows), we notice that for any particular representation, and any classifier, the accuracy obtained by additionally applying Group Lasso is higher, when compared to the same representation, and the same classifier used without Group Lasso. For example, when the representation is ST and the classifier is ANN, applying Group Lasso increases the performance by almost 4%. The improvement is even higher when the representation is GT, and the classifier is almost QDC. In this case the improvement is close to 33%.

Table 5: Performance of the classifiers without SLM and using Group Lasso for GT, MHWT, ST and combining all the dictionaries (GWST). The results by dictionaries and classifiers employed are indexed by rows and columns, respectively.

Dictionary	1-NN $\mu \pm \sigma$	3-NN $\mu \pm \sigma$	LDC $\mu \pm \sigma$	QDC $\mu \pm \sigma$	SVM $\mu \pm \sigma$	ANN $\mu \pm \sigma$
without SLM						
GT + Harmonics	0.7832 \pm 0.0129	0.7189 \pm 0.0156	0.4244 \pm 0.0175	0.6023 \pm 0.0165	0.7756 \pm 0.0165	0.6373 \pm 0.0489
MHWT + Harmonics	0.7614 \pm 0.0161	0.7484 \pm 0.0183	0.5743 \pm 0.0229	0.6637 \pm 0.0175	0.7514 \pm 0.0166	0.7579 \pm 0.0220
ST + Harmonics	0.9008 \pm 0.0152	0.8645 \pm 0.0171	0.6665 \pm 0.0209	0.8634 \pm 0.0111	0.8768 \pm 0.0364	0.9043 \pm 0.0503
GWST + Harmonics	0.8726 \pm 0.0143	0.8461 \pm 0.0136	0.6218 \pm 0.0244	0.8495 \pm 0.0158	0.8832 \pm 0.0188	0.8594 \pm 0.0542
Group Lasso						
GT + Harmonics	0.9283 \pm 0.0130	0.8949 \pm 0.0157	0.7416 \pm 0.0201	0.9318 \pm 0.0107	0.9008 \pm 0.0158	0.9261 \pm 0.0154
MHWT + Harmonics	0.8523 \pm 0.0184	0.8001 \pm 0.0198	0.6811 \pm 0.0185	0.8195 \pm 0.0142	0.8015 \pm 0.0355	0.8304 \pm 0.0338
ST + Harmonics	0.9629 \pm 0.0078	0.9402 \pm 0.0087	0.7069 \pm 0.0331	0.9566 \pm 0.0107	0.9422 \pm 0.0178	0.9358 \pm 0.0189
GWST + Harmonics	0.9752 \pm 0.0084	0.9560 \pm 0.0096	0.7589 \pm 0.0152	0.9593 \pm 0.0098	0.9588 \pm 0.0199	0.9598 \pm 0.0116

GT: Gabor transform. WT: Wavelet transform. ST: Stockwell transform. K-NN: K-nearest neighbours. LDC: linear discriminant classifier. QDC: quadratic discriminant classifier. SVM: support vector machine. ANN: artificial neural networks. μ and σ are the mean and the standard deviation of the PQ classification performance.

For all the classifiers, accuracy results between both ST and GWST, using Group Lasso, are similar. This is particularly true for QDC and SVM. Due to the similarity of these accuracies, we apply a Wilcoxon rank-sum test (WRS) to evaluate the statistical significance of both results per classifier, concluding that the differences in performance are statistically significant.

Tables 6 and 7 show the confusion matrices for GWST using 1-NN classifier without and with Group Lasso, respectively. From these tables we notice that without Group Lasso, misclassification occur in almost all the types of PQ disturbances, mainly in oscillatory distortions due to they require a high resolution in PQ representation both in time and frequency. On the other hand, the misclassification using Group Lasso are produced by the fact that similar frequency components appear in the different types of PQ disturbances (e.g. swell and sag, notch and oscillatory).

Table 6: Confusion matrix for GWST and a 1-NN classifier. Group Lasso is not applied in this case.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	46	5	2	0	0	0	4
C_2	8	46	0	0	0	0	3
C_3	2	0	52	0	1	0	2
C_4	0	0	0	57	0	0	0
C_5	0	0	0	0	57	0	0
C_6	0	0	0	0	0	56	1
C_7	2	1	2	2	4	7	39

Table 7: Confusion matrix for GWST and a 1-NN classifier. Group Lasso is applied in this case.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
C_1	57	0	0	0	0	0	0
C_2	0	56	1	0	0	0	0
C_3	0	2	55	0	0	0	0
C_4	0	0	0	57	0	0	0
C_5	0	0	0	0	57	0	0
C_6	0	0	0	0	0	57	0
C_7	0	0	0	0	2	0	55

C_1 : harmonic. C_2 : swell. C_3 : sag. C_4 : flicker. C_5 : notch. C_6 : impulsive transient. C_7 : oscillatory transient.

As an additional result, we evaluate the robustness of the framework proposed here in the presence of noise. In particular, we add white Gaussian noise to each PQ signal up until we get an specific Signal-to-Noise ratio (SNR). We then apply GWST plus Group Lasso over each signal, and run the classifiers twenty times, with different training and testing sets each time. Figure 3 shows the mean accuracy of the twenty repetitions obtained by each classifier, and for different SNR values. We notice that the classifiers show a reliable classification accuracy for SNR values greater than 30dB. For values below 30dB, the accuracy decreases due to the amount of noise, confusing ever more the waveforms among the PQ disturbance classification. Notice also that for low SNR values, more sophisticated classifiers such as ANN, QDC, and SVM behave better than simpler classifiers like LDC or those based on nearest neighbours.

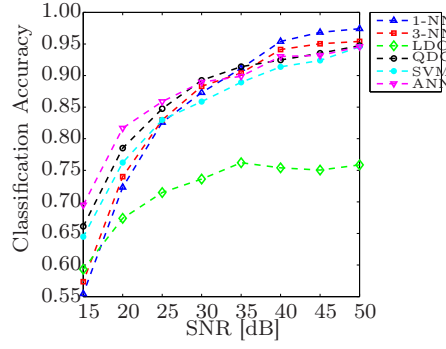


Figure 3: PQ disturbance classification performance using GWST plus Group Lasso under different levels of white Gaussian noise. NN: nearest neighbour. LDC: linear discriminant classifier. QDC: quadratic discriminant classifier. SVM: support vector machine. ANN: artificial neural network.

4 Conclusion

In this paper, we introduced overcomplete representations and sparse linear models for power quality disturbance classification. As an example of overcomplete representations, we combine the Gabor transform, the Wavelet transform with Mexican hat function, and the Stockwell transform, which are well known in the literature for PQ analysis. As an example of a sparse linear model, we use the Group Lasso assuming that each dictionary is a group in the sparse model.

Group Lasso selects automatically the most representative coefficients required to fully represent PQ disturbances, outperforming overcomplete representations. As we showed experimentally, this approach improves the performance of PQ disturbance classification for both linear and non-linear classifiers compared to methods without Group Lasso.

Acknowledgment

We are thankful to the Department of Electrical Engineering, Universidad Tecnológica de Pereira, for the technical support provided to carry out this research.

References

- [1] M. Bollen, I.-H. Gu, S. Santoso, M. McGranaghan, P. Crossley, M. Ribeiro, and P. Ribeiro, “Bridging the gap between signal and power,” *Signal Processing Magazine, IEEE*, vol. 26, no. 4, pp. 12–31, 2009.
- [2] S. Chattopadhyay, M. Mitra, and S. Sengupta, *Electric Power Quality*, 1st ed., ser. Power Systems. Springer, 2011.
- [3] A. Kusko and M. Thompson, *Power Quality in Electrical Systems*, 1st ed. McGraw-Hill Professional, 2007.
- [4] G. Benysek and M. Pasko, *Power theories for improved power quality*, ser. Power systems. Springer, 2012.
- [5] A. Farazmand, F. De Leon, K. Zhang, and S. Jazebi, “Analysis, modeling, and simulation of the phase-hop condition in transformers: The largest inrush currents,” *Power Delivery, IEEE Transactions on*, vol. 29, no. 4, pp. 1918–1926, 2014.
- [6] D. Granados, R. Romero, R. Osornio, A. Garcia, and E. Cabal, “Techniques and methodologies for power quality analysis and disturbances classification in power systems: a review,” *Generation, Transmission Distribution, IET*, vol. 5, no. 4, pp. 519–529, 2011.
- [7] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [8] G. K. Sharma, A. Kumar, C. B. Rao, T. Jayakumar, and B. Raj, “Short time Fourier transform analysis for understanding frequency dependent attenuation in austenitic stainless steel,” *NDT & E International*, vol. 53, pp. 1–7, 2013.

- [9] T. Vega, V. Roig, and H. San Segundo, "Evolution of signal processing techniques in power quality," in *Electrical Power Quality and Utilisation, 9th International Conference on*, 2007, pp. 1–5.
- [10] H. Eristi and Y. Demir, "Automatic classification of power quality events and disturbances using Wavelet transform and support vector machines," *Generation, Transmission Distribution, IET*, vol. 6, pp. 968–976, 2012.
- [11] C. Naik and P. Kundu, "Identification of short duration power quality disturbances employing S-transform," in *Power and Energy Systems (ICPS), 2011 International Conference on*, 2011, pp. 1–5.
- [12] N. Huang, L. Lin, W. Huang, and J. Qi, "Review of power-quality disturbance recognition using S-transform," in *Control, Automation and Systems Engineering (CASE), IITA International Conference on*, 2009, pp. 438–441.
- [13] N. M. Khoa, D. T. Viet, and N. H. Hieu, "Classification of power quality disturbances using Wavelet transform and K-nearest neighbor classifier," in *Industrial Electronics (ISIE), IEEE International Symposium on*, 2013, pp. 1–4.
- [14] J. Wang and C. Wang, "Bayes method of power quality disturbance classification," in *TENCON 2005 IEEE Region 10*, 2005, pp. 1–4.
- [15] P. G. V. Axelberg, I. Y. H. Gu, and M. H. J. Bollen, "Support vector machine for classification of voltage disturbances," *Power Delivery, IEEE Transactions on*, vol. 22, no. 3, pp. 1297–1303, 2007.
- [16] M. Reaz, F. Choong, M. Sulaiman, F. Mohd-Yasin, and M. Kamada, "Expert system for power quality disturbance classifier," *Power Delivery, IEEE Transactions on*, vol. 22, no. 3, pp. 1979–1988, 2007.
- [17] A. M. Gargoom, N. Ertugrul, and W. L. Soong, "Investigation of effective automatic recognition systems of power-quality events," *Power Delivery, IEEE Transactions on*, vol. 22, no. 4, pp. 2319–2326, 2007.
- [18] I. W. C. Lee and P. K. Dash, "S-transform-based intelligent system for classification of power quality disturbance signals," *Industrial Electronics, IEEE Transactions on*, vol. 50, no. 4, pp. 800–805, 2003.
- [19] T. Jayasree, D. Sam Harrison, and T. Sree Rangaraja, "Automated classification of power quality disturbances using Hilbert Huang transform and RBF networks," *International Journal of Soft Computing and Engineering*, vol. 1, no. 5, pp. 217–223, 2011.
- [20] D. Saxena and K. S. Verma, "Wavelet transform based power quality events classification using artificial neural network and SVM," *International Journal of Engineering, Science and Technology*, vol. 4, no. 1, pp. 87–96, 2012.
- [21] J. Kovacevic and A. Chebira, *An Introduction to Frames*, ser. Foundations and Trends in Signal Processing: Issue 1. Now Publishers, 2008, vol. 2, no. 1.
- [22] M. S. Lewicki, T. J. Sejnowski, and H. Hughes, "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337–365, 1998.
- [23] C. M. Bishop, *Pattern Recognition And Machine Learning (Information Science And Statistics)*. Springer, 2007.
- [24] K. P. Murphy, *Machine Learning: A Probabilistic Perspective (Adaptive Computation And Machine Learning Series)*. The MIT Press, 2012.
- [25] M. H. Bollen and I. Gu, *Signal processing of power quality disturbances*. Wiley-IEEE Press, 2006, vol. 30.
- [26] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [27] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 58, pp. 267–288, 1996.
- [28] L. Meier, S. van de Geer, and P. Bühlmann, "The Group Lasso for logistic regression," *Journal of the Royal Statistical Society*, vol. 70, no. 1, pp. 53–71, 2008.
- [29] M. Yuan, M. Yuan, Y. Lin, and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society*, vol. 68, pp. 49–67, 2006.
- [30] W. J. Fu, "Penalized Regressions: The Bridge versus the Lasso," *Journal of Computational and Graphical Statistics*, vol. 7, no. 3, pp. 397–416, 1998.
- [31] P. Janik and T. Lobos, "Automated classification of power-quality disturbances using SVM and RBF networks," *Power Delivery, IEEE Transactions on*, vol. 21, no. 3, pp. 1663–1669, 2006.

- [32] T. Abdel-Galil, M. Kamel, A. Youssef, E. El-Saadany, and M. Salama, “Power quality disturbance classification using the inductive inference approach,” *Power Delivery, IEEE Transactions on*, vol. 19, no. 4, pp. 1812–1818, 2004.
- [33] S. Roy and S. Nath, “Classification of power quality disturbances using features of signals,” *International Journal of Scientific and Research Publications*, vol. 2, no. 11, pp. 1–9, 2012.
- [34] I. Monedero, C. Leon, J. Ropero, A. Garcia, J. Elena, and J. C. Montano, “Classification of electrical disturbances in real time using neural networks,” *Power Delivery, IEEE Transactions on*, vol. 22, no. 3, pp. 1288–1296, 2007.